Advanced Regression: Overview of the course

Garyfallos Konstantinoudis

Feb 20, 2024

Course aims

- Understand the fundamental concepts of **data science** including high-dimensional statistics and advanced regression
- Translate these into **real world data problems**:
  - **Understanding** structure in complex data
  - **Prediction** of future observations
- **Evaluate** results of other studies and publications
- **Design** and conduct analysis of complex datasets

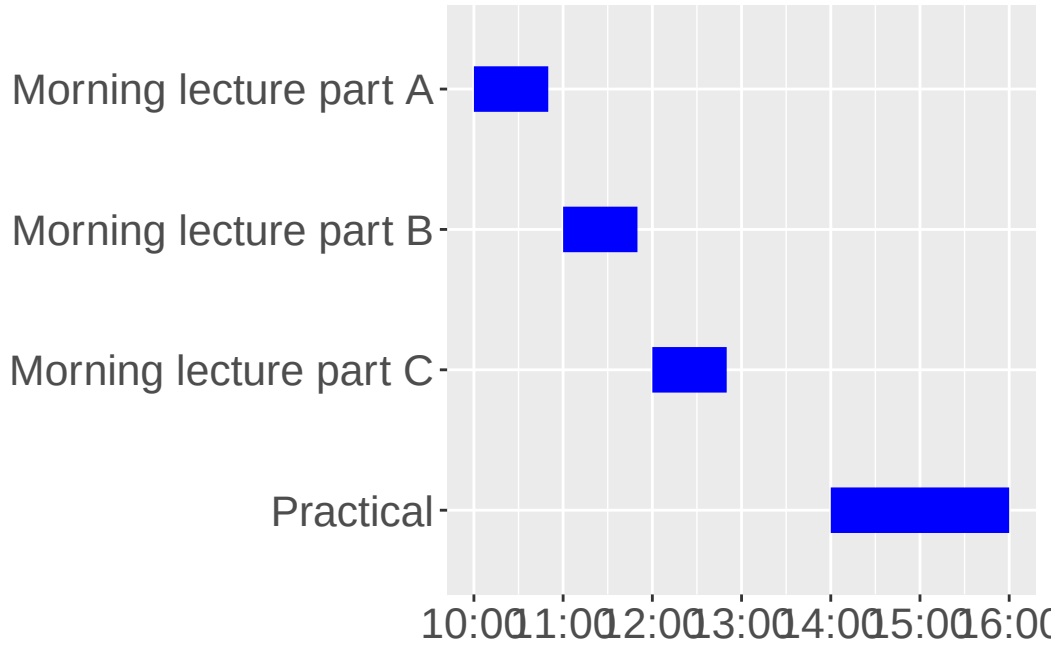
Topics covered

1. **Understanding** structure in complex data:
   - Linear and generalised linear models
   - Hierarchical models
   - Non-linear exposures
   - Model and variable selection
2. **Prediction** of future observations:
   - Prediction rules and cross validation
   - Penalised regression
   - Machine learning (classification, trees, neural networks)

Module structure

```
Warning: package 'ggplot2' was built under R version 4.3.2
```

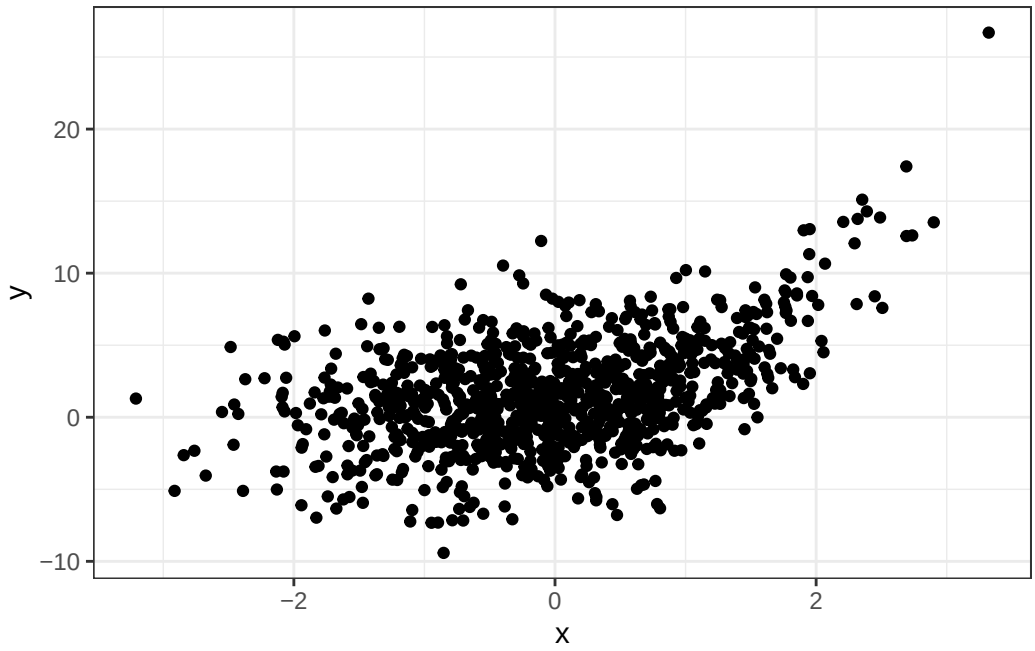
```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```



Practicals in R

There will be an R practical following each course:

```
x <- rnorm(n = 1000, mean = 0, sd = 1)
y <- exp(x) + rnorm(n = 1000, mean = 0, sd = 3)
ggplot2::ggplot() +
  ggplot2::geom_point(data = data.frame(x = x, y = y), ggplot2::aes(x = x, y = y)) +
  ggplot2::theme_bw()
```



Practicals

- Opportunity to interact with module lead and tutors.
- Discussion of the solutions.
- Forum on padlet for support.
- Solutions will be released after the practical session.
- Working on the practicals and solving the questions is the best preparation for the exam.

What is markdown or quarto?

- When coding in R it is important to document and comment the code.
- Markdown is an R package that compiles R code into documents (pdf, html, word and many more).
- https://cran.r-project.org/web/packages/rmarkdown.
- Project page https://rmarkdown.rstudio.com/.

Make your code accessible and reproducible.

Learning outcomes

- Perform advanced statistical analyses, employing penalised likelihood or non-pararametric regression models.
- Discuss the theoretical foundations and limitations of the most widely used advanced regression approaches.
- Use complex regression models in R, understand which methods are suitable for which data and interpret the results.
- Enjoy data science.

Drop in sessions

- Wednesdays (11:00-12:00).
- In person and/or online.
- Opportunity to re-discuss and get feedback.

Exam

2 exams:

1. An interim exam (30%) during the 3rd week. Not in person, analyse data based on lectures 1 and 2 and 24h deadline: Wednesday 06.03 at 09:00.
2. A final 1.5h open book exam during the exam period (70%).

The total mark and pass will be calculated after **combining** these two marks.

Slides format

- All the material are available on Blackboard and Github
- https://github.com/gkonstantinoudis/advanced-regression
- Check Introduction to Git course at data camp (not mandatory!)
- Material is available as pdf and html. To get the html format as the lectures, download the .qmd files and render the document. Make sure you have
  - All required packages
  - Rtools
  - LaTeX

AR: The team

- Module lead: Garyfallos Konstantinoudis
- Practical facilitators:
  - TBC

Looking forward to welcoming you to the AR module!

Any questions?

- g.konstantinoudis@imperial.ac.uk

Next lectures

Linear models and generalised linear models.

- Revision: The linear model.
- Generalised linear model.

Random effects models.

- Motivation: Structured data.
- Fixed and random effects.